Assignment 1
Assigned on Sep 14th, 2021. Due date Sep 27th, 2021 before midnight.

Instructions

(1) Please read the instruction carefully. Also take this habit with you into the exam room.
(2) Please read each question carefully and answer the questions straightforwardly. Always provide economic reasons at lease a paragraph for your analysis, or a graph when necessary, even when the question does not indicate so.
(3) Handing and submitting assignments are only available via BE Moodle.

Answering the questions and preparing answer sheets

(1) Answers are to be handwritten, in either digital or analog form, in a blank canvas or any clean paper. Make sure that your handwriting is clearly visible and readable.
(2) There is no need to rewrite the question. Just indicate the question number clearly for each of the answer, such as 1.a).
(3) Default decimal point is 4.
(4) Choose precise wordings, especially when you want to interpret the meaning of a test, confidence interval, or coefficients.
(5) When done, for the digital case, collage all the pages into a single PDF file. For those who write on sheets of paper, take photo of all pages then convert all of them into a single PDF file as well.
(6) Name your PDF file as StudentID_YourNickname, such as 640123456_Bo.

Submitting your answers

(1) Make sure your file does not exceed 10MB. This is the maximum file size for BE Moodle upload.
(2) Login to BE Moodle, head into the course, then the assignment topic.
(3) Choose your file to submit. Done. There will be timestamp for your upload date and time, so please make sure to not submit later than that.

Assignment 1
Assigned on Sep 14th, 2021. Due date Sep 27th, 2021 before midnight.

---

1. **(15 points)** Given this information,

$n = 46$	$\sum_{i=1}^{n} X_i = 3{,}959.80$	$\sum_{i=1}^{n} Y_i = 3{,}180.80$
$\bar{X} = 86.0826$	$\bar{Y} = 69.1478$	
$\sum_{i=1}^{n}(X_i)^2 = 364{,}023.30$		$\sum_{i=1}^{n} X_i Y_i = 319{,}943.18$
$\sum_{i=1}^{n}(X_i - \bar{X})^2 = 23{,}153.3861$		$\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = 94{,}525.1748$
$\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}) = 46{,}131.6183$		$\sum_{i=1}^{n} \hat{u}_i^{\,2} = 2{,}610.9211$

answer the following questions. Show your work.

a) (4 points) From regression model: $Y_i = \beta_1 + \beta_2 X_i + u_i$, $u_i \sim NIID\left(0, \sigma^2\right)$, find the estimators of β_1 and β_2 with OLS method and explain the meaning of the model.

b) (2 points) Find R^2 and explain its meaning.

c) (1 points) If $X_i = 60$, estimate the value of $\hat{Y}_i$ and explain its meaning.

d) (3 points) Find the estimators of $\text{var}\left(u_i\right)$, $\text{var}\left(\hat{\beta}_1\right)$ and $\text{var}\left(\hat{\beta}_2\right)$

e) (2.5 points) What are the 95-percent confident intervals for β_2? Interpret the meaning.

f) (2.5 points) Test the hypothesis whether coefficients (both β_1 and β_2) are different from zero at 0.05 level of significance.

1

a) (4 points) From regression model: $Y_i = \beta_1 + \beta_2 X_i + u_i$, $u_i \sim NIID\left(0, \sigma^2\right)$, find the estimators of β_1 $\hat{\beta}_1$ and β_2 $\hat{\beta}_2$ with OLS method and explain the meaning of the model.

$$\hat{\beta}_2 = \frac{\sum x_i (Y_i - \bar{Y})}{\sum x_i (x_i - \bar{x})}$$

$$= \frac{\sum (x_i - \bar{x})(Y_i - \bar{Y})}{\sum (x_i - \bar{x})^2}$$

$$= \frac{46{,}131.6183}{23{,}153.3461}$$

$$= 1.9924$$

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}$$

$$= 69.1478 - 1.9924(86.0826)$$

$$= -102.3632$$

$$\therefore \quad Y_i = \beta_1 + \beta_2 X_i + u_i \Rightarrow \hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$$

$$Y_i = -102.3632 + 1.9924 X_i$$

At -102.3632 and slope is 1.9924. When X_i increase by 1 unit, $\hat{Y}_i$ increase by 1.9924

b) (2 points) Find R^2 and explain its meaning.

$$r^2 = 1 - \frac{\sum \hat{u}_i^2}{\sum (Y_i - \bar{Y})^2}$$

$$= 1 - \frac{2{,}610.9219}{94{,}525.1748}$$

$$= 0.9724$$

$$= 97.24\%$$

$\therefore$ r^2 values of 0.9724 suggets that X_i explain about 97.24% of the variation in Y_i

c) (1 points) If $X_i = 60$, estimate the value of $\ddot{Y}_i$ and explain its meaning.

From $Y_i = \beta_1 + \beta_2 X_i + u_i$

$Y_i = -102.9632 + 1.9924(60)$

$Y_i = 17.1808$; When X_i is 60, Y_i is 17.1808

$\therefore$ We are able to replace any valve out of sample and estimate Y_i but there is an error.

d) (3 points) Find the estimators of $\operatorname{var}(u_i)$, $\operatorname{var}(\hat{\beta}_1)$ and $\operatorname{var}(\hat{\beta}_2)$

$var(u_i)$

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-k}$$

$$= \frac{2,610.9211}{46-2}$$

$$= 59.3391$$

$var(\hat{\beta}_1)$

$$\hat{\sigma}^2_{\hat{\beta}_1} = \frac{\sum X_i^2}{n\sum x_i^2}\left(\frac{\sum \hat{u}_i^2}{n-k}\right)$$

$$= \frac{\sum X_i^2}{n\sum (X_i-\bar{X})^2}\left(\frac{\sum \hat{u}_i^2}{n-k}\right)$$

$$= \frac{314,023.9}{(46)(23,153.3817)}\left(\frac{2,610.9211}{46-2}\right)$$

$$= 20.2994$$

$var(\hat{\beta}_2)$

$$\hat{\sigma}^2_{\hat{\beta}_2} = \left(\frac{\sum \hat{u}_i^2}{n-k}\right)\left(\frac{1}{\sum x_i^2}\right)$$

$$= \left(\frac{\sum \hat{u}_i^2}{n-k}\right)\left(\frac{1}{\sum (X_i-\bar{X})^2}\right)$$

$$= \left(\frac{2610.9211}{46-2}\right)\left(\frac{1}{23,153.3817}\right)$$

$$= 0.00256 \ ?$$

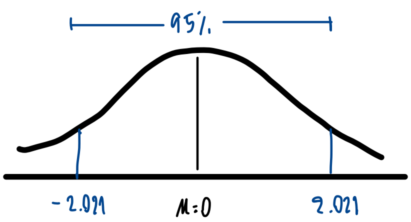
e) (2.5 points) What are the 95-percent confident intervals for β_2? Interpret the meaning.

$\alpha = 0.05$

$t_{\frac{\alpha}{2}} = t_{\frac{0.05}{2}} = t_{0.025}$

$df = 46-2$

$t_{0.025}, df = 44 = \pm 2.021$

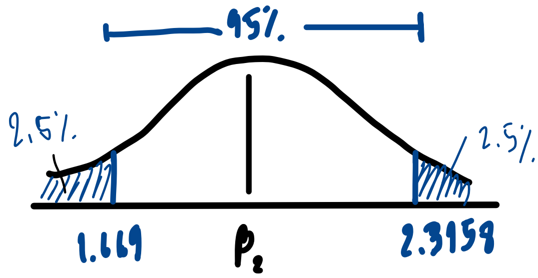


$\hat{\beta}_2 \pm t_{\frac{\alpha}{2}} \cdot \hat{\sigma}_{\hat{\beta}_2}$

upper : $1.9924 + (2.021 \times \sqrt{0.0256}) = 2.3158$

lower : $1.9924 - (2.021 \times \sqrt{0.0256}) = 1.669$

$\therefore P[1.669 \leq \beta_2 \leq 2.3158] = 0.95$



อธิบาย interpret

f) (2.5 points) Test the hypothesis whether coefficients (both β_1 and β_2) are different from zero at 0.05 level of significance. $\alpha = 0.05$

$\underline{\beta_1}$

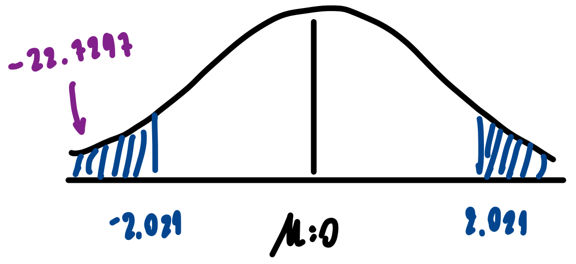
$H_0 : \beta_1 = 0$ → Null Hypothesis

$H_a : \beta_1 \neq 0$

$$t_{cal} = \frac{\hat{\beta}_1 - \beta_1}{\sigma\hat{\beta}_1} = \frac{-102.3637 - 0}{4.5035} = -22.7297$$

$$t_{\frac{\alpha}{2}} = t_{\frac{0.05}{2}} = t_{0.025}, \quad df = 46-2 = 44$$

$$t_{\frac{\alpha}{2}} = \pm 2.021$$



→ $t_{cal} = -22.7297$ falls into rejection area. So, we reject H_0 (Null Hypothesis)

∴ We are 95% confidence sure that β_1 is not equal to 0.

$\underline{\beta_2}$

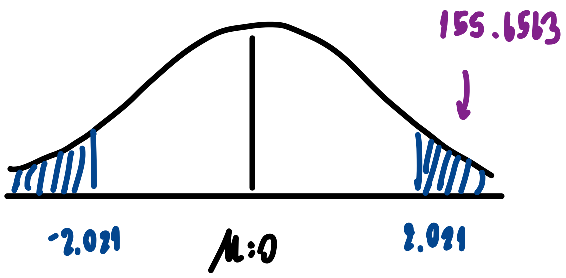
$H_0 : \beta_1 = 0$ → Null Hypothesis

$H_a : \beta_2 \neq 0$

$$t_{cal} = \frac{\hat{\beta}_2 - \beta_2}{\sigma\hat{\beta}_2} = \frac{1.9924 - 0}{0.0128} = 155.6536$$

$$t_{\frac{\alpha}{2}} = t_{\frac{0.05}{2}} = t_{0.025}, \quad df = 46-2 = 44$$

$$t_{\frac{\alpha}{2}} = \pm 2.021$$



→ $t_{cal} = 155.6563$ falls into rejection area. So, we reject H_0 (Null Hypothesis)

∴ With 95% confidence, we are sure that β_2 is not equal to 0.

2. **(8 points)** Answer the following question without any mathematical proof. A 3-6-line paragraph for a question is sufficient.
   a) (2 points) If we have only one data point, can we create a sample regression function? Why?
   b) (2 points) Does a significant β_2 sufficient for us to believe that X and Y are causally related? Provide an example to support your answer.
   c) (2 points) When we test a hypothesis and find that β_2 is significantly different from zero, what does the result actually suggest? Answer with specific wording from statistical perspective.
   d) (2 points) What is (are) (an) advantage(s) of an interval estimation over point estimate?

3. **(7 points)** Given that the dependent variable is natural log of wage (lwage) in Thai Baht and the independent variable is hours worked per week (main_hr), the result of estimation is shown in the table below here.

Source	SS	df	MS			
				Number of obs	=	308
				F(1, 306)	=	92.20
ESS Model	50.060869	1	50.060869	Prob > F	=	0.0000
RSS Residual	166.152715	306	.542982728	R-squared	=	0.2315
				Adj R-squared	=	0.2290
TSS Total	216.213584	307	.704278775	Root MSE	=	.73687

					lower limit	upper limit
(Dependent) Y lwage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
(Independent) X main_hr	$(\hat{\beta}_1)$ .0318017	.003312 $\sigma_{\hat{\beta}_1}$	9.60	0.000	.0252844	.0383189
_cons	$(\hat{\beta}_0)$ 7.658082	.1256392 $\sigma_{\hat{\beta}_0}$	60.95	0.000	7.410856	7.905308

Answer the following questions. Show your work.
   a) (2 points) On average, how much is the nominal wage for a person who works 0 hour a week? (Note that this is a point estimation, not a prediction)
   b) (2 points) If a person works an hour more, how much, on average, wage change do we expect?
   c) (3 points) If you want to change the reading of unit from hours worked to days worked, what values in the main_hr row will differ? Calculate the changes to all the values in that row, **disregarding the constant row**. You might want to impose an assumption here. State that clearly before calculation.

****

2

a) (2 points) If we have only one data point, can we create a sample regression function? Why?

Ans : We are able to create a sample regression function by estimate the values which are β_1 and β_2 are called parameters, where as $\hat{\beta}_1$ and $\hat{\beta}_2$ are called estimators in order to calculate the slope and intercept.

b) (2 points) Does a significant β_2 sufficient for us to believe that X and Y are causally related?

Ans : By using statistical significand test, if the value of tcall falls into rejection area we can reject the Null-Hypothesis which can imply that β_2 is not zero. In another word, X and Y are related.
For example, the equation $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$ and the other variables are detemined in the table below, using OLS method

$n = 46$	$\sum_{i=1}^{n} X_i = 3{,}959.80$	$\sum_{i=1}^{n} Y_i = 3{,}180.80$
$\bar{X} = 86.0826$	$\bar{Y} = 69.1478$	
$\sum_{i=1}^{n}(X_i)^2 = 364{,}023.30$		$\sum_{i=1}^{n} X_i Y_i = 319{,}943.18$
$\sum_{i=1}^{n}(X_i - \bar{X})^2 = 23{,}153.3861$		$\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = 94{,}525.1748$
$\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}) = 46{,}131.6183$		$\sum_{i=1}^{n} \hat{u}_i^{\,2} = 2{,}610.9211$

We found that $\hat{\beta}_1 = -102.3632$ and $\hat{\beta}_2 = 1.9924$ written as $\hat{Y}_i = -102.3632 + 1.9924 X_i$. The slope is equal to $1.9924 X_i$ which is β_2 implying that when X_i increase by 1 unit $\hat{Y}_i$ increase by 1.9924 Therefore, a significant β_2 is sufficient to believe the relations of X and Y.

c) (2 points) When we test a hypothesis and find that β_2 is significantly different from zero, what does the result actually suggest? Answer with specific wording from statistical perspective.

Ans: If β_2 is significantly different from zero, in other word, the t_{cal} falls into rejection area. So, we reject H_0 (Null-Hypothesis) or reject that $H_0 : \beta_2 = 0$. The type I error presents how confidence in term of percentage you are sure that β_2 is not equal to 0.

d) (2 points) What is (are) (an) advantage(s) of an interval estimation over point estimate?

Ans: When we select one point in the sample to represents a point in the population is called point estimation. This estimation can show the exact value of the population point estimation.

It enables to tell the confidence that the interval will conclude the true population value, but it cannot express the overall estimated population value as the interval estimation.

3

	Source	SS	df	MS
ESS	Model	50.060869	1	50.060869
RSS	Residual	166.152715	306	.542982728
TSS	Total	216.213584	307	.704278775

Number of obs	=	308
F(1, 306)	=	92.20
Prob > F	=	0.0000
R-squared	=	0.2315
Adj R-squared	=	0.2290
Root MSE	=	.73687

	lwage		Coef.	Std. Err.	t	P>|t|	[95% Conf. (lower limit)	Interval] (upper limit)
(dependent) Y	lwage							
(Independent) X	main_hr	$(\hat{\beta}_1)$	.0318017	.003312 $\sigma_{\hat{\beta}_1}$	9.60	0.000	.0252844	.0383189
	_cons	$(\hat{\beta}_0)$	7.658082	.1256392 $\sigma_{\hat{\beta}_0}$	60.95	0.000	7.410856	7.905308

a) (2 points) On average, how much is the nominal wage for a person who works 0 hour a week? (Note that this is a point estimation, not a prediction)

$$lwage = 7.658082 + 0.031807(Main\_hr)$$

$$lwage = 7.658082 + 0.031807(0)$$

$$lwage = 7.6581$$

$$wage = e^{7.6581}$$

$\therefore$ Nominal wage = 2117.73 for a person who work 0 hour.

\#

b) (2 points) If a person works an hour more, how much, on average, wage change do we expect?

According to $lwage = 7.6581 + 0.0318(Main\_hr)$,

$$\frac{d\,lwage}{d\,Main\_hr} = 0.0318$$

$$\frac{d\,\widehat{wage}}{wage} = 0.0318(Main\_hr)$$

$$\%\Delta\widehat{wage}_i = 0.0318(Main\_hr) = 3.18\%$$

$\Rightarrow$ When the person work more for an hour, we expect wages to increase by 3.18%

c) (3 points) If you want to change the reading of unit from hours worked to days worked, what values in the main_hr row will differ? Calculate the changes to all the values in that row, **disregarding the constant row**. You might want to impose an assumption here. State that clearly before calculation.

Change from hours worked per week to days worked per week, we need to divide by 24 according to a day has 24 hours.

From the equation $\widehat{lwage} = 7.6581 + 0.0318\,(Main\_hr)$

When $Y = lwage$
$X = Main\_hr$
$\hat{\beta}_2 = 0.0318$
$SE = (0.1256)(0.003312)$

Put X divided by 24, we will get
$\hat{\beta}_2$ is $\hat{\beta}_2 \cdot 24 \rightarrow 0.0318(24) = 0.7632$
$SE(\hat{\beta}_2)$ is $SE(\hat{\beta}_2) \cdot 24 \rightarrow 0.003312(24) = 0.0795$

The new equation will be $\widehat{lwage} = 7.6581 + 0.7632\,(Main\_days)$
$SE = (0.1256)\,(0.0795)$

$\therefore$ Therefore, the values that will change in the main_hr row are the coefficient and the standard Error (SE) which the coefficient will change from 0.0318 to 0.7632 and the standard error will change from 0.00312 to 0.0795 instead.